

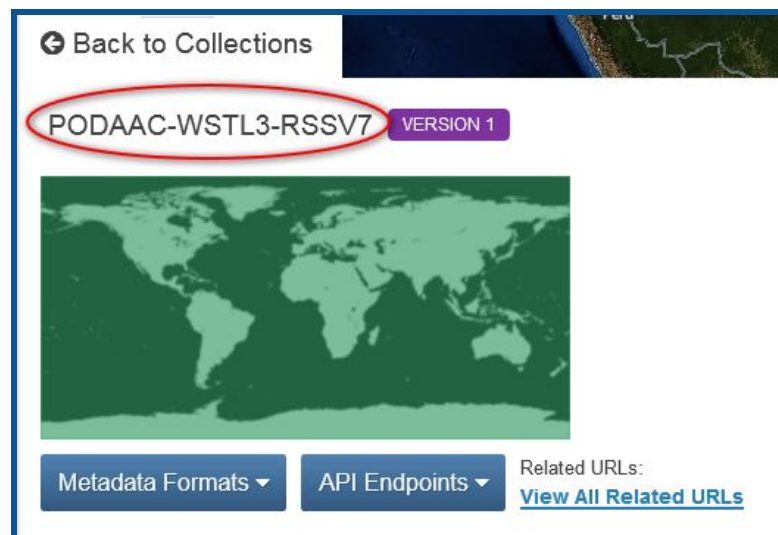
# Summary of ARC's PODAAC Metadata Findings

*July 5, 2017*

This report outlines the ARC team's metadata findings for PODAAC. The findings reported below are present in the majority of the collection level and granule level metadata records. Detailed reports for individual collection and granule level metadata records are provided separately. The detailed reports will be used for tracking the DAAC's progress towards working off the reported findings and metrics will be generated from the detailed reports. An explanation of the color coding found in the detailed reports is included in this report along with preliminary metrics.

## 1. Collection Level Findings

- i. 'VersionId' in the metadata is '1' across most of PODAAC's collection level metadata records. However, the version provided in the title of the dataset or in supporting documentation is different. Please either update the version element to include the current version number or confirm that '1' is the correct version for the dataset.
- ii. Currently the Data Set Id is presented in an abbreviated format (for example: PODAAC-AQR40-3WRCS). Because 'DataSetId' maps to 'EntryTitle' in CMR, the DataSetId should be a descriptive title of the dataset. Therefore, ARC recommends changing the information provided in the DataSetId element to a descriptive title of the dataset. If possible, this title should match the title of the dataset associated with the DOI. PODAAC can consider switching the 'LongName' and 'DataSetId' fields in the metadata, or formulate another method to fulfill this recommendation.



- iii. For the dataset description element, many metadata records only provide the long name of the dataset as the description or only provide a description which is one sentence long. The 'Description' element is intended to provide a summary of the dataset, mimicking a journal abstract that is useful to the science community but also approachable for a first-time user of the data. ARC recommends replacing the short descriptions with the abstract provided on the dataset landing page.
- iv. Whenever possible, the long name or title should be included in the first sentence of the description. Providing the title in the first sentence of the description provides clarity for a user who discovers the dataset via the CMR API or has downloaded the metadata directly. The DOI title should be used, when applicable, for consistency.
- v. A new DOI element has been added to the ECHO10 metadata schema (<https://git.earthdata.nasa.gov/projects/EMFD/repos/echo-schemas/browse/schemas/10.0/Collection.xsd#226>). DOI is a required element in CMR for NASA datasets. The DOI element is where the DOI string should be provided for all collection records. Here is a sample of how DOI should be provided in ECHO10 metadata:  

```
<DOI>  
    <DOI>10.5067/LIS/LIS-OTD/DATA308</DOI>  
</DOI>
```
- vi. The 'ProcessingCenter' field should follow GCMD vocabulary: <https://gcmdservices.gsfc.nasa.gov/static/kms/providers/providers.csv>
- vii. The 'ArchiveCenter' element is also a GCMD controlled vocabulary field (<https://gcmdservices.gsfc.nasa.gov/static/kms/providers/providers.csv>). Please change "PO.DAAC" to "NASA/JPL/PODAAC" in order to be compliant with the GCMD vocabulary. PODAAC should contact GCMD directly if a change in the naming convention is desired.
- viii. ARC recommends providing the dataset citation in the 'CitationForExternalPublication' element. The citation provided on the dataset landing page should be leveraged whenever possible. Based on recent discussions at the metadata summits, it is likely that the 'CitationForExternalPublication' element in ECHO10 will be broken into several separate fields, rather than be a single block of text.
- ix. The 'Collection Progress' element is now a required element in UMM and is meant to describe the production status of the collection. There are three responses for the 'Collection Progress' element and those responses are chosen from a controlled vocabulary list. Valid responses are: PLANNED, IN WORK and COMPLETE. Note that 'Collection Progress' is

the equivalent of the ECHO10 element called 'Collection State.' Also note that the valid responses for this element are subject to change during the next UMM-C version update.

- x. The Temporal 'RangeDateTime' information should be as accurate as possible by matching the temporal information in the first and last granule. The beginning date and time should match the start date and time of the first granule in the collection. Similarly, the ending date and time should match the stop date and time of the final granule in the collection.
- xi. Currently PODAAC metadata lists a <ContactPerson> with first name "User," middlename "null" and last name "Services." There is also a <JobPosition> provided with the value of "Technical Contact." ARC recommends removing the <ContactPerson> elements since a <OrganizationName> is already provided.
- xii. The Platform/Type element is a controlled vocabulary element and should adhere to the 'Category' portion of the GCMD Platforms/Sources keyword list ([https://gcmdservices.gsfc.nasa.gov/static/kms/platforms/platforms.csv?ed\\_wiki\\_keywords\\_page](https://gcmdservices.gsfc.nasa.gov/static/kms/platforms/platforms.csv?ed_wiki_keywords_page))
- xiii. Whenever the 'Sensor' information is a repetition of the 'Instrument' information, ARC recommends that the 'Sensor' information be removed from the metadata.
- xiv. The campaign short name (Campaign/ Short name) information should be added to all collection level metadata records. This element is required in CMR and leverages the GCMD 'Projects' keyword list ([https://gcmdservices.gsfc.nasa.gov/static/kms/projects/projects.csv?ed\\_wiki\\_keywords\\_page](https://gcmdservices.gsfc.nasa.gov/static/kms/projects/projects.csv?ed_wiki_keywords_page)). If a campaign or project is not included in the GCMD keyword list, a request can be made to the GCMD team to include the project. Finally, if there is a campaign long name, it should be included in the Campaign/Long Name element as well.
- xv. PODAAC currently provides many data access links via FTP. Data access is preferably provided via HTTPS instead of FTP. ARC recommends the migration of data access to HTTPS whenever possible in order to encourage DAAC alignment with the NASA requirement.
- xvi. OPeNDAP links should be moved to the OnlineResource element in ECHO 10 so that these links can be assigned an appropriate URL type and subtype.
- xvii. All OnlineResourceURLs should have a URL Description. The description should be unique to each URL. If a new URL is added to the metadata, the URL should have a description.

- xviii. OnlineResource Types should be compliant with the UMM-Common content type keyword list. Accepted URL content type lists can be found in the [UMM-Common documentation](#) and the UMM-Common schema ([Type enum](#), [Subtype enum](#)). The CMR API attempts to map the online resource type included in the native metadata to UMM-Common content types. If online resource types are not compliant, translation errors can occur. Leveraging the UMM-Common keyword list prevents these translation errors from occurring and also ensures a consistent user experience in the Earthdata search client. Note that in ECHO 10 there is no breakdown of type or subtype, therefore a value can be selected from either the type or subtype list.
- xix. All OPeNDAP links and THREDDS links should include a mime type. Mime type values should be provided for all services and should be selected from the values provided in the [UMM-Common documentation](#) (page 27). Mime type values can also be found in the [UMM-Common schema](#). ARC and the CMR team can provide guidance on mime types as needed.
- xx. Many of PODAAC's datasets have spatial reference information in the user's guide or other documentation. Whenever possible, ARC strongly recommends including information about the datum and geographic coordinate system associated with the data in the metadata that is provided to CMR.

## **2. Granule Level Findings**

- i. Data Set ID (when included in the granule metadata) is presented in an abbreviated format (for example: PODAAC-AQR40-3WRCS). Because 'DataSetId' maps to 'EntryTitle' in CMR, the DataSetId should be a descriptive title of the dataset. Therefore, ARC recommends changing the information provided in the DataSetId element to a descriptive title of the dataset. If possible, this title should match the title of the dataset associated with the DOI. PODAAC can consider switching the 'LongName' and 'DataSetId' fields in the metadata, or formulate another method to fulfill this recommendation.
- ii. 'VersionId' in the metadata is '1' across most of PODAAC's collection level metadata records. However, the version provided in the title of the dataset or in supporting documentation is different. Please either update the version element to include the current version number or confirm that '1' is the correct version for the dataset.

- iii. Many values provided in the 'DataGranule/SizeMBDataGranule' are provided in KB rather than MB. If possible, please convert the granule size to MB to eliminate confusion for users.
- iv. Consider providing the platform short name, instrument short name, and campaign short name in the granule level metadata.
- v. Currently, many of the Online Access URLs point to a FTP link. Access to data should be migrated from ftp to https to comply with EOSDIS policy. URS authentication should be implemented in front of data access.
- vi. All OPeNDAP links should include a mime type. Per UMM-Common, all service links are required to have a mime type. Valid mime type values can be found on page 27 of the UMM-Common documentation ([https://wiki.earthdata.nasa.gov/download/attachments/49448405/UMM-Common\\_20161219.doc?version=2&modificationDate=1482237063584&api=v2](https://wiki.earthdata.nasa.gov/download/attachments/49448405/UMM-Common_20161219.doc?version=2&modificationDate=1482237063584&api=v2)). Mime type values can also be found in the [UMM-Common schema](#).
- vii. ARC recommends providing a description for all Online Access URLs and all Online Resource URLs. Descriptions should be unique to each URL.
- viii. ARC recommends adding the DOI link to landing page to the granule level metadata whenever possible.
- ix. If any 'OnlineResource' links are added to the granule level metadata, the URL Type must be provided and should be compliant with the UMM-Common content type keyword list. Accepted URL content type lists can be found in the [UMM-Common documentation](#) and the UMM-Common schema ([Type enum](#), [Subtype enum](#)). Note that in ECHO 10 there is no breakdown of type or subtype, therefore a value can be selected from either the type or subtype list.

### 3. Explanation of Color Coding Provided in Detailed Reports

Color	Definition
Cyan	Required field based on UMM-C
Light Purple	An optional primary element with required sub-elements based on UMM-C
Purple	A sub-element which is only required if any information is provided in the scope of the primary element based on UMM-C

White	Completely optional field
Red	Correcting these issues should be of the highest priority
Yellow	Correcting these errors are strongly recommended but are not required
Blue	Minor error/ inconsistency; points out features noticed by the ARC Team which may help improve the robustness of the metadata but are not required to be addressed
*	Any field with an asterisk is controlled by GCMD vocabulary

#### 4. Metrics

269 Collection level records checked

<i>Collection Level</i>	# Red fields	# Yellow fields	# Blue fields	Total # fields checked
	4,381	1,196	1,093	14,924
	29.4%	8.0%	7.3%	

229 Granule level records checked

<i>Granule Level</i>	# Red fields	# Yellow fields	# Blue fields	Total # fields checked
	547	485	1,819	6,438
	8.5%	7.5%	28.3%	

498 Total records checked (collection + granule)

<i>Cumulative</i>	# Red fields	# Yellow fields	# Blue fields	Total # fields checked
	4,928	1,681	2,912	21,362
	23.1%	7.9%	13.6%	

## 5. Miscellaneous Dataset Issues

1. The following datasets have DOI links that are not working. DOI URLs were found on the [PO.DAAC DOI wiki page](#):
  - a. GRACE WEEKLY NON-TIDAL ATMOSPHERE GEOPOTENTIAL COEFFICIENTS GFZ RELEASE 5.0 (GRACE\_GAA\_L2\_GRAV\_GFZ\_RLo5\_WEEK) - <https://cmr.earthdata.nasa.gov/search/concepts/C1293258563-PODAAC/1.umm-json>
  - b. GRACE WEEKLY NON-TIDAL OCEAN GEOPOTENTIAL COEFFICIENTS GFZ RELEASE 5.0 (GRACE\_GAB\_L2\_GRAV\_GFZ\_RLo5\_WEEK) - <https://cmr.earthdata.nasa.gov/search/concepts/C1293259391-PODAAC/1.umm-json>
  - c. GRACE WEEKLY NON-TIDAL ATMOSPHERE AND OCEAN GEOPOTENTIAL COEFFICIENTS GFZ RELEASE 5.0 (GRACE\_GAC\_L2\_GRAV\_GFZ\_RLo5\_WEEK) - <https://cmr.earthdata.nasa.gov/search/concepts/C1293260405-PODAAC/1.umm-json>
  - d. GRACE WEEKLY OCEAN BOTTOM GEOPOTENTIAL COEFFICIENTS GFZ RELEASE 5.0 (GRACE\_GAD\_L2\_GRAV\_GFZ\_RLo5\_WEEK) - <https://cmr.earthdata.nasa.gov/search/concepts/C1293261338-PODAAC/1.umm-json>
  - e. MASCON CLMR4 Scale Factor with CRI Filter (TELLUS\_MASCON\_CLM4\_CRI\_SCALE\_V1) - <https://cmr.earthdata.nasa.gov/search/concepts/C1293893190-PODAAC/1.umm-json>
  - f. MASCON Land Mask used with CRI filter (TELLUS\_MASCON\_LAND\_MASK\_CRI\_V1) - <https://cmr.earthdata.nasa.gov/search/concepts/C1293893225-PODAAC/1.umm-json>
  - g. Gridded Altimeter Fields with Enhanced Coastal Coverage Daily (ALT\_TIDE\_GAUGE\_L4\_OST\_SLA\_US\_WEST\_COAST\_DAILY) -

<https://cmr.earthdata.nasa.gov/search/concepts/C1293166439-PODAAC/1.umm-json>

2. The following datasets cannot be found via the PODAAC search tool. No granules are available for these datasets in CMR and the OPeNDAP links are not functioning properly. Data is still available via ftp:
  - a. Jason-1 Microwave Radiometer version E (JASON-1\_JMR\_E) -  
<https://cmr.earthdata.nasa.gov/search/concepts/C1291532213-PODAAC/1.umm-json>
  - b. Jason-1 Microwave Radiometer version E Geodetic (JASON-1\_JMR\_E\_GEODETTIC) -  
<https://cmr.earthdata.nasa.gov/search/concepts/C1291955234-PODAAC/1.umm-json>
3. The following datasets could not be found in CMR:
  - a. CCMP\_RSS\_L3.o\_V2.o
  - b. RSCAT\_LEVEL\_2B\_OVV\_CLIM\_12\_V2
4. The following datasets have new versions:
  - a. Integrated Multi-Mission Ocean Altimeter Data for Climate Research complete time series Version 3 (MERGED\_TP\_J1\_OSTM\_OST\_ALL\_V3) -  
<https://cmr.earthdata.nasa.gov/search/concepts/C1293272468-PODAAC/1.umm-json>. There is a new version of this data set (version 4). The Online Access URL for this record links to version 4 of the data. The links at the granule level metadata, which link to version 3 of the data set don't work. Version 4 metadata is not yet in CMR.
  - b. Integrated Multi-Mission Ocean Altimeter Data for Climate Research Version 3 (MERGED\_TP\_J1\_OSTM\_OST\_CYCLES\_V3) -  
<https://cmr.earthdata.nasa.gov/search/concepts/C1293274042-PODAAC/1.umm-json> There is a new version of this data set (version 4). The Online Access URL for this record links to version 4 of the data. The links at the granule level metadata, which link to version 3 of the data set don't work. Version 4 metadata is not yet in CMR.